特征加权支持向量机

汪廷华 田盛丰 黄厚宽
(北京交通大学计算机与信息技术学院 北京 100044)

摘 要：该文针对现有的加权支持向量机(WSVM)和模糊支持向量机(FSVM)只考虑样本重要性而没有考虑特征重要性对分类结果的影响的缺陷，提出了基于特征加权的支持向量机方法，即特征加权支持向量机(FWSVM)。该方法首先利用信息增益计算各个特征对分类任务的重要度，然后用获得的特征重要度对核函数中的内积和欧氏距离进行加权计算，从而避免了核函数的计算被一些弱相关或不相关的特征所支配。理论分析和数值实验的结果都表明，该方法比传统的 SVM 具有更好的鲁棒性和分类能力。

关键词：支持向量机；特征加权；信息增益；机器学习
中图分类号：TP18 **文献标识码**： A **文章编号**：1009-5896(2009)03-0514-05

Feature Weighted Support Vector Machine

Wang Ting-hua Tian Sheng-feng Huang Hou-kuan
(*School of Computer and Information Technology*, *Beijing Jiaotong University*, *Beijing* 100044, *China*)

Abstract: Support vector machine has been applied in many research fields, such as pattern recognition and function estimate. There is a shortcoming in Weighted SVM and Fuzzy SVM, which take the importance of sample into account but neglect the relative importance of each feature with respect to the classification task. In this paper a SVM approach is proposed based on the feature weighting, i.e. Feature Weighted SVM (FWSVM). This method first estimates the relative importance (weight) of each feature by computing the information gain. Then, it utilizes the weights for computing the inner product and Euclidean distance in kernel functions. In this way the computing of kernel function can avoid being dominated by trivial relevant or irrelevant features. Theoretical analysis and experimental results show that the FWSVM is more robust and has the better performance of generalization than the traditional SVM.

Key words: Support Vector Machine (SVM); Feature weighting; Information gain; Machine learning

1 引言

支持向量机(Support Vector Machine, SVM)是在Vapnik等人所建立的统计学习理论(Statistical Learning Theory, STL)基础上发展起来的一种新的学习算法，它基于VC维理论和结构风险最小化原理，根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折衷，在很大程度上克服了传统机器学习中的维数灾难和局部极小等问题，从而获得较好的泛化能力[1,2]。正是基于上述原因，这一方法已经广泛应用于模式识别和函数估计等问题中。为了改善支持向量机的分类性能，许多研究人员提出了改进的方法。Lin和张翔等将模糊数学引入支持向量机，提出了模糊支持向量机(Fuzzy SVM, FSVM)模型[3,4]，它根据不同输入样本对分类贡献的不同，赋以相应的隶属度，这样可以减少野值和噪声的影响，提高分类性能。为了克服不同类别样本尺寸大小相差较为悬殊而导致分类性能不尽人意的缺点，范昕炜等提出了对每一类样本赋以不同权值的加权支持向量机(Weighted SVM, WSVM)模型[5]，对类别差异造成的影响进行了相应的补偿，从而提高了小类别样本的分类精度，这对于某些需要重点关注小类别样本精度的应用研究有重要的现实意义。赵晖等综合WSVM和FSVM的思想，提出了一种同时考虑了类别样本数差异和样本重要性差异的双重加权支持向量机(Dual-Weighted SVM, DWSVM)模型[6]，进一步改善了分类性能。

然而，上述方法都是从样本重要程度的角度来考虑问题，而没有从特征重要程度的角度来分析问题；它们都隐含假定待分析样本向量的各维特征对分类的贡献均匀，不考虑各个特征对分类的不同影响。核函数是SVM研究的核心问题，常用核函数是通过描述样本相似性的内积或距离来定义的，而内积或距离是根据样本的所有特征计算的。在这些特征中，有些特征与分类强相关，有些特征与分类弱相关，还有些特征与分类不相关。这样，核函数的计算可能会被一些弱相关或不相关的特征所支配，从而影响分类器的分类性能。

基于上述问题的分析，本文提出一种基于特征加权的支持向量机方法，简称特征加权支持向量机(Feature Weighted SVM, FWSVM)。FWSVM首先用信息增益评估相应于分类

2007-10-31 收到，2008-04-07 改回
国家 973 规划项目(2007CB307100，2007CB307106)和北京交通大学科技基金(2006XM007)资助课题

任务的各个特征的重要性，并依据重要性给各个特征设定相应的权重，使强相关特征获得比弱相关特征更大的权重。然后将获得的权重应用到核函数的计算中，从而避免核函数的计算被一些弱相关或不相关的特征所支配。数值实验结果表明，FWSVM比传统的SVM具有更好的泛化能力和鲁棒性。

2　支持向量机

支持向量机是从线性可分情况下的最佳超平面发展而来的。对于一组带有类别标记的训练样本集$(\boldsymbol{x}_i,y_i),\boldsymbol{x}_i\in R^n$, $y_i\in\{+1,-1\},i=1,\cdots,l$，若超平面$\tilde{\boldsymbol{w}}\cdot\boldsymbol{x}+b=0$能将样本正确分为两类，则最佳超平面应使两类样本到超平面最小距离之和最大。最佳超平面通过求解下面的优化问题得到

$$\left.\begin{aligned}&\min\ \frac{1}{2}\|\tilde{\boldsymbol{w}}\|^2+C\sum_{i=1}^{l}\xi_i\\&\text{s.t.}\quad y_i[(\tilde{\boldsymbol{w}}\cdot\boldsymbol{x}_i)+b]\geq 1,\ i=1,\cdots,l\\&\qquad\xi_i\geq 0,\ i=1,\cdots,l\end{aligned}\right\}\tag{1}$$

其中ξ_i为误差项，C为惩罚系数。

利用Lagrange乘子法可以把上述求解最佳超平面的问题转化为对偶问题

$$\left.\begin{aligned}&\max\ \sum_{i=1}^{l}a_i-\frac{1}{2}\sum_{i,j=1}^{l}a_ia_jy_iy_j(\boldsymbol{x}_i\cdot\boldsymbol{x}_j)\\&\text{s.t.}\quad\sum_{i=1}^{l}a_iy_i=0,\\&\qquad 0\leq a_i\leq C,\quad i=1,\cdots,l\end{aligned}\right\}\tag{2}$$

其中a_i是样本点$\boldsymbol{x}_i$的Lagrange乘子。这是一个典型的二次函数寻优问题，存在唯一解。可以证明解中只有一部分(通常是少部分)a_i不为零，对应的样本就是支持向量。由此得到决策函数

$$f(\boldsymbol{x})=\text{sign}\left(\sum_{i=1}^{l}a_iy_i(\boldsymbol{x}_i\cdot\boldsymbol{x})+b\right)\tag{3}$$

对于线性不可分情形，先利用一个非线性映射$\phi:R^n\rightarrow H$将数据映射到一个高维的特征空间H，在特征空间中可以实现线性分类。通过定义函数$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\phi(\boldsymbol{x}_i)\cdot\phi(\boldsymbol{x}_j)$，式(2)的对偶规划变为

$$\left.\begin{aligned}&\max\ \sum_{i=1}^{l}a_i-\frac{1}{2}\sum_{i,j=1}^{l}a_ia_jy_iy_jk(\boldsymbol{x}_i,\boldsymbol{x}_j)\\&\text{s.t.}\quad\sum_{i=1}^{l}a_iy_i=0,\\&\qquad 0\leq a_i\leq C,\quad i=1,\cdots,l\end{aligned}\right\}\tag{4}$$

相应的决策函数变为

$$f(\boldsymbol{x})=\text{sign}\left(\sum_{i=1}^{l}a_iy_ik(\boldsymbol{x}_i,\boldsymbol{x})+b\right)\tag{5}$$

其中$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\phi(\boldsymbol{x}_i)\cdot\phi(\boldsymbol{x}_j)$称为核函数，核函数的选取应使其为特征空间的一个内积。常用的核函数有：

(1)多项式核函数

$$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=(\boldsymbol{x}_i\cdot\boldsymbol{x}_j+1)^d,\ d=1,2,\cdots\tag{6}$$

(2)Gauss 径向基核函数

$$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\exp(-\gamma\,\|\,\boldsymbol{x}_i-\boldsymbol{x}_j\,\|^2)\tag{7}$$

(3)Sigmoid 核函数

$$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\tanh(b(\boldsymbol{x}_i\cdot\boldsymbol{x}_j)+c),\quad b>0,\ c<0\tag{8}$$

3　特征加权

依据某种准则对数据集中的各个特征赋予一定的权重称为特征加权。应用特征加权提高机器学习算法性能的研究可参见文献[7–9]，这些文献均是用特征权重向量$\boldsymbol{w}$扩充标准的欧氏距离

$$d^w(\boldsymbol{x}_i,\boldsymbol{x}_j)=\sqrt{\sum_{k=1}^{n}w_k\,|\,x_{ik}-x_{jk}\,|^2}\tag{9}$$

其中$d^w(\boldsymbol{x}_i,\boldsymbol{x}_j)$表示两个样本$\boldsymbol{x}_i$和$\boldsymbol{x}_j$的加权欧氏距离，$x_{ik}$表示样本$\boldsymbol{x}_i$的第$k$个特征的值，$\boldsymbol{w}=(w_1,\cdots,w_n)$是权重向量，$w_k\geq 0(k=1,\cdots,n)$是相应于各个特征的重要性权重。显然，$\boldsymbol{w}\neq(1,\cdots,1)$意味着欧氏空间中的轴将依据$w_k$的大小而拉伸或收缩：较大$w_k$对应的轴将收缩，而较小$w_k$对应的轴将拉伸；换句话说，就是由于引入$\boldsymbol{w}$而导致了欧氏空间中点与点之间位置关系的改变。

在特征加权中$\boldsymbol{w}$的求取是关键。特征权重的计算是特征(属性)相关分析的重要内容。在分类学习中特征相关分析的基本思想是计算某种度量，用于量化特征与给定类别的相关性。这里介绍常用的基于信息增益的特征相关分析[10,11]。

设D是具有类标号特征的样本(元组)数据集。假定类标号特征具有m个不同的值，定义m个不同的类$C_i(i=1,\cdots,m)$。设$C_{i,D}$是D中类C_i的样本的集合，$|\,D\,|$和$|\,C_{i,D}\,|$分别是D和$C_{i,D}$中样本的个数。则对D中的样本分类所需的期望信息由下式给出

$$\text{Info}(D)=-\sum_{i=1}^{m}p_i\log_2(p_i)\tag{10}$$

其中p_i是任意样本属于C_i的概率，并用$|\,C_{i,D}\,|/|\,D\,|$估计。现在假设要按特征A(非类标号)划分D中的样本，其中特征A根据训练数据的观测具有v个不同值$\{a_1,a_2,\cdots,a_v\}$，可以用A将D划分为v个子集$\{D_1,D_2,\cdots D_v\}$，其中$D_j(j\in\{1,2,\cdots,v\})$包含$D$中这样一些样本，它们在$A$上具有值$a_j$。根据$A$的这种划分对$D$的样本分类所需要的期望信息由下式给出

$$\text{Info}_A(D)=\sum_{j=1}^{v}\frac{|\,D_j\,|}{|\,D\,|}\times\text{Info}(D_j)\tag{11}$$

信息增益定义为原来的信息需求(即仅基于类比例)与新的需求(即对A划分之后得到的)之间的差，即

$$\text{Gain}(A)=\text{Info}(D)-\text{Info}_A(D)\tag{12}$$

在这种方法中，可以计算样本集D中每个特征(非类标号)的信息增益，具有最高信息增益的特征是给定特征集合中具有最高区分度的特征，亦即对分类贡献最大的特征。由此可以用信息增益来度量各个特征相对于分类的相关性：信息

增益越大，相关性越强。假设数据集 D 中的每个样本(类标号除外)由 n 个特征 $(f_1,\cdots,f_n)$ 来描述，则向量 $\boldsymbol{G}=(\mathrm{Gain}(f_1),\cdots,\mathrm{Gain}(f_n))$ 自然地描述了各个特征的权重，其中 $\mathrm{Gain}(f_i)$ 表示特征 f_i 的信息增益。

4 构造特征加权支持向量机

基于特征加权核函数构造的支持向量机称为特征加权支持向量机。这里先给出特征加权核函数的定义。

定义1 令 k 是定义在 $X\times X$ 上的核函数，$X\subseteq R^n$，$\boldsymbol{P}$ 是给定输入空间的 n 阶线性变换矩阵，其中 n 是输入空间的维数。特征加权核函数 $k_{\boldsymbol{P}}$ 定义为

$$k_{\boldsymbol{P}}(\boldsymbol{x}_i,\boldsymbol{x}_j)=k(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{P},\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{P}) \tag{13}$$

线性变换矩阵 $\boldsymbol{P}$ 也被称为特征加权矩阵，矩阵 $\boldsymbol{P}$ 的不同选择导致不同的加权情形：

(1) $\boldsymbol{P}$ 是 n 阶单位阵。此时就是没有加权的情形。

(2) $\boldsymbol{P}$ 是 n 阶对角阵。此时 $(\boldsymbol{P})_{ii}=w_i\ (1\le i\le n)$ 代表第 i 个特征的权重，一般来说 w_i 不全相等。

$$\boldsymbol{P}=\begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{bmatrix}$$

(3) $\boldsymbol{P}$ 是任意 n 阶方阵。此时称之为全加权的情形。

$$\boldsymbol{P}=\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix}$$

本文只考虑 $\boldsymbol{P}$ 是对角阵的情形。

定义 2 由式(13)得到下列常用的特征加权核函数：

(1)特征加权多项式核函数

$$k_{\boldsymbol{P}}(\boldsymbol{x}_i,\boldsymbol{x}_j)=(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{P}\cdot\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{P}+1)^d=(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{P}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{x}_j+1)^d,\quad d=1,2,\cdots \tag{14}$$

(2)特征加权 Gauss 径向基核函数

$$\begin{aligned} k_{\boldsymbol{P}}(\boldsymbol{x}_i,\boldsymbol{x}_j)&=\exp(-\gamma\ ||\ \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{P}-\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{P}\ ||^2)\\ &=\exp(-\gamma((\boldsymbol{x}_i-\boldsymbol{x}_j)^{\mathrm{T}}\boldsymbol{P}\boldsymbol{P}^{\mathrm{T}}(\boldsymbol{x}_i-\boldsymbol{x}_j))) \end{aligned} \tag{15}$$

(3)特征加权 Sigmoid 核函数

$$\begin{aligned} k_{\boldsymbol{P}}(\boldsymbol{x}_i,\boldsymbol{x}_j)&=\tanh(b(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{P}\cdot\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{P})+c)\\ &=\tanh(b(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{P}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{x}_j)+c),\quad b>0,\ c<0 \end{aligned} \tag{16}$$

引入核函数的原始动机是利用非线性特征映射建立的特征空间中的线性函数寻找非线性模式。矩阵$\boldsymbol{P}$看起来可能和这个动机不相关，因为它相当于线性特征映射。但是这种映射在实际中很有用，因为它不仅可以缩放输入空间的几何形状，还可以缩放特征空间的几何形状，从而改变分配给特征空间中不同线性函数的权重。定理1表述了这一结论。

定理 1 令 k 是定义在 $X\times X$ 上的核函数，$X\subseteq R^n$，ϕ 是输入空间到特征空间的映射 $\phi:X\to F$，$\boldsymbol{P}$是线性变换矩阵，$\tilde{\boldsymbol{x}}_i=\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{P}$，则 $\left\|\phi(\tilde{\boldsymbol{x}}_i)-\phi(\tilde{\boldsymbol{x}}_j)\right\|\ne\left\|\phi(\boldsymbol{x}_i)-\phi(\boldsymbol{x}_j)\right\|$。

证明 对于Gauss径向基核函数(或规格化多项式核函数、规格化Sigmoid核函数)，有

$k(\boldsymbol{x},\boldsymbol{x})=1,\ \forall\boldsymbol{x}$ 成立。则

$$\begin{aligned} \left\|\phi(\tilde{\boldsymbol{x}}_i)-\phi(\tilde{\boldsymbol{x}}_j)\right\|&=\sqrt{(\phi(\tilde{\boldsymbol{x}}_i)-\phi(\tilde{\boldsymbol{x}}_j))\cdot(\phi(\tilde{\boldsymbol{x}}_i)-\phi(\tilde{\boldsymbol{x}}_j))}\\ &=\sqrt{\phi(\tilde{\boldsymbol{x}}_i)\cdot\phi(\tilde{\boldsymbol{x}}_i)-2\phi(\tilde{\boldsymbol{x}}_i)\cdot\phi(\tilde{\boldsymbol{x}}_j)+\phi(\tilde{\boldsymbol{x}}_j)\cdot\phi(\tilde{\boldsymbol{x}}_j)}\\ &=\sqrt{k(\tilde{\boldsymbol{x}}_i,\tilde{\boldsymbol{x}}_i)-2k(\tilde{\boldsymbol{x}}_i,\tilde{\boldsymbol{x}}_j)+k(\tilde{\boldsymbol{x}}_j,\tilde{\boldsymbol{x}}_j)}\\ &=\sqrt{2-2k(\tilde{\boldsymbol{x}}_i,\tilde{\boldsymbol{x}}_j)}=\sqrt{2-2k(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{P},\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{P})}\\ &\ne\sqrt{2-2k(\boldsymbol{x}_i,\boldsymbol{x}_j)}=\left\|\phi(\boldsymbol{x}_i)-\phi(\boldsymbol{x}_j)\right\| \end{aligned}$$

证毕

定理 2 若存在一个 $w_k=0$，$1\le k\le n$，则数据集的第 k 个特征与加权核函数的计算无关，与分类器的输出无关。$w_k\ (1\le k\le n)$越小，对加权核函数的计算影响越小，对分类结果的影响越小。

证明 由加权核函数的定义和分类决策函数式(5)立即得证。 证毕

定理1说明经过线性特征变换之后，特征空间的形状改变(特征空间中点与点之间位置关系改变)了，从而有可能在改变之后的特征空间中找到更好的线性分类超平面，提高SVM的分类性能。定理2说明加权核函数的计算能够避免被一些弱相关或不相关的特征所支配，从而期望获得更好的分类结果。本文后面的数值实验肯定了这个结论。

特征加权支持向量机(FWSVM)的构造步骤如下：

步骤1 收集数据集 S，S 中的样本由特征集合 $\langle\mathrm{fs},d\rangle$ 描述，其中 d 为类标号特征(分类特征)，$\mathrm{fs}=(f_1,f_2,\cdots,f_n)$ 为非类标号特征集合。

步骤2 计算特征 $f_i(1\le i\le n)$ 的信息增益，构造特征权重向量 $\boldsymbol{w}$ 和线性变换矩阵 $\boldsymbol{P}$ (对角阵)：

$\boldsymbol{w}=\sqrt{\boldsymbol{G}}=\left(\sqrt{\mathrm{Gain}(f_1)},\cdots,\sqrt{\mathrm{Gain}(f_n)}\right)$，$\boldsymbol{P}=\mathrm{diag}(\boldsymbol{w})$，其中 $(\boldsymbol{P})_{kk'}=w_k\delta_{kk'}$。

步骤3 用特征加权核函数式(14)-式(16)替换标准SVM中核函数式(6)-式(8)，选用合适的模型与训练算法对数据集 S 构建分类决策函数(分类器)。

步骤4 对获得的分类器进行性能评估。

5 实验结果与分析

为了验证所提出的方法，从UCI机器学习数据库[12]中选择了名为“Wisconsin Breast Cancer”和“Letter-recognition”两个数据库做了实验。前者按年代顺序共收集了八组数据，本文将这八组数据合并成一组数据，总共有699个样本，每个样本由10个特征描述(其中一个是类标号)，除去其中16个包含未知特征值的样本，剩下683个样本，记为 S_1，S_1 是二分类数据集。后者是字母图像识别数据库，总共有20000个样本，每个样本由17个特征描述(其中一个是类标号)，记为 S_2，S_2 是多分类数据集(总共26类，分别表示26个英文字母)。为了比较FWSVM和C-SVM泛化能力的差异，分别在 S_1

和 S_2 中随机抽取样本总数的1/8作为训练集，剩下的7/8作为测试集。由于测试样本数远远大于训练样本数，测试精度可以看作是学习算法真实泛化能力的体现。

本文设计了基于LIBSVM软件包[13]的C++程序，开发环境为Visual C++6.0，运行环境为Windows 2003 server。实验中选用Gauss径向基核函数式(7)和特征加权Gauss径向基核函数式(15)。为了使实验更具说服力，选取了5组不同的SVM参数(C,γ) ($0<C<100$, $0.01<\gamma<1$)值进行了比较，(C,γ)值如表1所示。在数据集 S_1 上的实验结果如表2所示，在数据集 S_2 上的实验结果如表3所示。

表1 5组SVM参数

组号	1	2	3	4	5
C	70	2	56	35	93
γ	0.1	0.05	0.5	0.8	0.08

表2 数据集S_1上的实验结果

组号	C-SVM		FWSVM	
	测试精度(%)	支持向量个数	测试精度(%)	支持向量个数
1	94.8161	56	95.4849	49
2	95.6522	50	96.6555	41
3	83.9465	72	88.9632	61
4	81.6054	73	84.9498	70
5	95.1505	54	95.4849	46
平均	90.2341	61	92.3077	53

表3 数据集S_2上的实验结果

组号	C-SVM		FWSVM	
	测试精度(%)	支持向量个数	测试精度(%)	支持向量个数
1	87.4629	2259	91.3371	1793
2	89.6514	2096	90.5314	1716
3	43.1200	2465	76.3657	2300
4	25.1371	2469	61.8000	2386
5	88.5943	2204	91.4286	1753
平均	66.7931	2299	82.2926	1990

从实验结果可以看出，FWSVM在选定的数据集上的泛化能力优于标准的C-SVM：对数据集 S_1 来说，使用FWSVM使平均测试精度增加了2.3%，支持向量减少了13.11%；对数据集 S_2 来说，使平均测试精度增加了23.21%，支持向量减少了13.44%。

FWSVM在数据集 S_2 上获得了比 S_1 更好的结果的原因可以从学习算法的鲁棒性上加以解释。在实验中我们获得了 S_1 和 S_2 的特征权重向量并计算它们的数字特征：

$\boldsymbol{w}_{s1}$ =[0.6812, 0.8380, 0.8227, 0.6815, 0.7310, 0.7766, 0.7452, 0.6980, 0.4604]；

$\boldsymbol{w}_{s2}$ = [0.3274, 0.1453, 0.4029, 0.2317, 0.3828, 0.6981, 0.9018, 0.8170, 0.8521, 0.7697, 0.9298, 0.8671, 0.9523, 0.8949, 0.8726, 0.6283]。

$\boldsymbol{w}_{s1}$ 和 $\boldsymbol{w}_{s2}$ 的数字特征比较如表4所示，数据集 S_1 和 S_2 的特征权重分布如图1和图2所示。

表4 $\boldsymbol{w}_{s1}$ 和 $\boldsymbol{w}_{s2}$ 的数字特征比较

	最大值	最小值	平均值	标准偏差
$\boldsymbol{w}_{s1}$	0.8380	0.4604	0.7150	0.1113
$\boldsymbol{w}_{s2}$	0.9523	0.1453	0.6671	0.2753

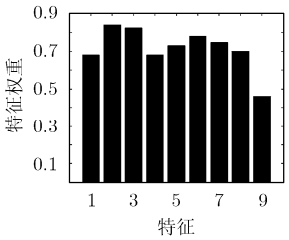

图1 数据集S_1的特征权重分布图

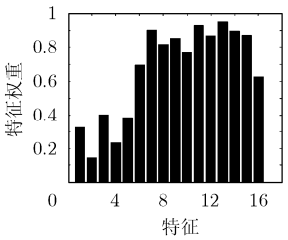

图2 数据集S_2的特征权重分布图

由表4，图1与图2可以看出：$\boldsymbol{w}_{s2}$ 的波动程度大于 $\boldsymbol{w}_{s1}$ 。换句话说，数据集 S_2 中的不同特征对分类结果影响程度的差别较 S_1 大，而FWSVM正是利用特征加权核函数反映这种差别的学习算法，它通过特征加权减少弱相关特征(权重小)对分类结果的影响，从而提高了学习算法的鲁棒性。

6 结束语

本文在对样本加权型支持向量机(如WSVM和FSVM)分析的基础上，提出了特征加权型的支持向量机，即FWSVM。讨论了特征加权核函数的构造及其性质。理论分析表明通过特征加权不仅可以改变输入空间中点与点之间的位置关系，还可以改变特征空间中点与点之间的位置关系；特征加权核函数的计算能够避免被一些弱相关或不相关的特征所支配。基于信息增益的数值实验结果表明，本文提出的方法能够比较明显地提高分类器的泛化能力。显然，如何将样本加权型支持向量机和特征加权型支持向量机集成起来，进一步提高支持向量机的分类性能是需要我们进一步研究的问题。

参 考 文 献

[1] Vapnik V. The Nature of Statistical Learning Theory. New York: SpringerVerlag, 1995: 91-188.

[2] Cristianini N and Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge: Cambridge University Press, 2000: 47-98.

[3] Lin C F and Wang S D. Fuzzy support vector machines. *IEEE Trans. on Neural Networks*, 2002, 13(2): 464-471.

[4] 张翔, 肖小玲, 徐光右. 基于样本之间紧密度的模糊支持向量机方法. 软件学报, 2006, 17(5): 951–958.
Zhang Xiang, Xiao Xiao-ling, and Xu Guang-you. Fuzzy support vector machines based on affinity among samples. *Journal of Software*, 2006, 17(5): 951–958.

[5] 范昕炜, 杜树新, 吴铁军. 可补偿类别差异的加权支持向量机算法. 中国图像图形学报, 2003, 8(9): 1037–1042.
Fan Xin-wei, Du Shu-xin, and Wu Tie-jun. Weighted support vector machine based classification algorithm for uneven class size problems. *Chinese Journal of Image and Graphics*, 2003, 8(9): 1037–1042.

[6] 赵晖, 荣莉莉. 支持向量机组合分类及其在文本分类中的应用. 小型微型计算机系统, 2005, 26(10): 1816–1820.
Zhao Hui and Rong Li-li. Classification algorithm based on combined support vector machines and its application in text categorization. *Mini-Micro Systems*, 2005, 26(10): 1816–1820.

[7] Zhan Yan, Chen Hao, and Hang Guochun. An optimization algorithm of K-NN classifier. Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian, China, 2006: 2246–2251.

[8] Wang Xizhao, Wang Yadong, and Wang Lijuan. Improving fuzzy c-means clustering based on feature-weight learning. *Pattern Recognition Letters*, 2004, 25(10): 1123–1132.

[9] 李洁, 高新波, 焦李成. 基于特征加权的模糊聚类新算法. 电子学报, 2006, 34(1): 89–92.
Li Jie, Gao Xin-bo, and Jiao Li-cheng. A new feature weighted fuzzy clustering algorithm. *Acta Electronica Sinica*, 2006, 34(1): 89–92.

[10] Quinlan J R. Induction of decision tree. *Machine Learning*, 1986, 1(1): 81–106.

[11] Han Jiawei and Kamber M. Data Mining: Concepts and Techniques, Second Edition. Beijing: China Machine Press, 2006: 296–300.

[12] Asuncion A and Newman D J. UCI Machine Learning Repository. http://www.ics.uci.edu/~mlearn/ MLRepository.html, 2007.

[13] Chang Chihchung and Lin Chihjen. LIBSVM: a library for support vector machines. http://www.csie.ntu.tw/~cjlin/libsvm, 2007.

汪廷华： 男，1977 年生，博士生，研究方向为机器学习、数据挖掘.

田盛丰： 男，1944 年生，教授，博士生导师，主要研究方向为人工智能、模式识别、网络安全等.

黄厚宽： 男，1940 年生，教授，博士生导师，主要研究方向为人工智能、KDD、多 Agent 等.